NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

NAVY ENLISTMENT SUPPLY MODEL AT THE RECRUITING STATION LEVEL

by

Claude M. McRoberts

June 2008

Thesis Advisor: Samuel E. Buttrey
Second Reader: David L. Schiffman

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2008	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE Navy Enlistment Supply Model at the Recruiting Station Level	**5. FUNDING NUMBERS**
6. AUTHOR(S) Claude M. McRoberts	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Past econometric studies have sought insight into the factors that affect military enlistment supply by creating models based on econometric theory and testing them with data in order to confirm their proposed theoretical relationships. The purpose of this study is to utilize factors common to previous research along with the additional factors of proximity to military installations and high school quality to build the best predictive model. This study utilizes data from 2002 through 2006 to predict high-quality male active-duty Navy enlistments at the recruiting station level. This study shows that neural network models tend to predict the best, followed by regression-based models and then tree-based models. The number of recruiters assigned per Navy Recruiting Station (NRS) and the male 17- to 19-year-old populations proved to be the most important predictive factors. The number of houses, veteran population percentage, land area, percentage of high school students receiving subsidized lunches, Navy installation proximity and per capita were common to all predictive models. This study also finds that NRSs closer to larger navy installations, having higher student-to-teacher ratios, having lower graduation rates (measured by "Promoting Power") and having lower percentages of students on subsidized lunches exhibit greater high-quality male enlistment rates.

14. SUBJECT TERMS Recruiting, Enlistment Supply, Regression, Tree, Neural Network, Military, Navy, Manpower, Prediction, Navy	**15. NUMBER OF PAGES** 57
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

NAVY ENLISTMENT SUPPLY MODEL AT THE RECRUITING STATION LEVEL

Claude M. McRoberts
Lieutenant Commander, United States Navy
B.A., Indiana State University, 1998

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2008

Author: Claude M. McRoberts

Approved by: Samuel E. Buttrey
Thesis Advisor

David L. Schiffman
Second Reader

James N. Eagle
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Past econometric studies have sought insight into the factors that affect military enlistment supply by creating models based on econometric theory and testing them with data in order to confirm their proposed theoretical relationships. The purpose of this study is to utilize factors common to previous research along with the additional factors of proximity to military installations and high school quality to build the best predictive model. This study utilizes data from 2002 through 2006 to predict high-quality male active-duty Navy enlistments at the recruiting station level. This study shows that neural network models tend to predict the best, followed by regression-based models and then tree-based models. The number of recruiters assigned per Navy Recruiting Station (NRS) and the male 17- to 19-year-old populations proved to be the most important predictive factors. The number of houses, veteran population percentage, land area, percentage of high school students receiving subsidized lunches, Navy installation proximity and per capita were common to all predictive models. This study also finds that NRSs closer to larger navy installations, having higher student-to-teacher ratios, having lower graduation rates (measured by “Promoting Power”) and having lower percentages of students on subsidized lunches exhibit greater high-quality male enlistment rates.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The purpose of this study was to build a model that accurately predicts the number of high-quality male Navy enlistments at the Navy Recruiting Station (NRS) level. This study also sought to explore the relationship between military installation proximity and high-quality male Navy enlistments and between various measures of public high school quality and high-quality male Navy enlistments.

This study aggregated zip code and county level data from several different sources to the NRS level. The number of males with Armed Forces Qualification Test (AFQT) scores at 50 or above who joined the Navy's delayed entry program (DEP) as determined from Defense Manpower Data Center (DMDC) data was used as the response variable. Population data provided by Woods & Poole Economics, veteran population data derived from the 2000 Census, the number of recruiters per NRS supplied by Navy Recruiting Command (CNRC), unemployment data downloaded from the Department of Labor, income data gathered from the Department of Commerce, and public high school data retrieved from the Department of Education was used to develop models and relationships in this study.

Through the use of regression trees, ordinary least squares multiple linear regression models, and neural networks, the study concluded that NRSs closer to larger navy installations produced higher numbers of high-quality male enlistments. Additionally, NRSs whose territories have higher student-to-teacher ratios, lower "Promoting Power" scores (a measure of high school graduation rates), and lower percentages of students on subsidized lunches produce greater numbers of high-quality male enlistments rates. This study also concluded that neural network models outperform both regression and tree models in predicting high-quality male Navy enlistments at the NRS level.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

> In the 21st Century, our most sophisticated weapon system is the human brain, and our most powerful advantage is our people. Today, and in future operations, people provide the margin of performance that determines who wins or loses, succeeds or fails, in pursuit of vital national interests.[1]
>
> — Department of the Navy, *Human Capital Strategy 2007*

As stated in the quotation, people are the most important component of the United States Navy. In order for the Navy to achieve the level of performance necessary to succeed, it must bring in the best people in sufficient numbers. Navy Recruiting Command (CNRC) is responsible for recruiting the necessary 51,997 people into the Navy for Fiscal Year (FY) 2008.[2]

CNRC employs nearly 7,200 military and civilian personnel dispersed throughout the United States and abroad to fill the ranks of the Navy. CNRC has organized these personnel into a single headquarters, two regions, twenty-six districts (NRDs), and over 1,500 stations (NRSs). Territory is uniquely assigned by zip code to each NRS.[3]

Since CNRC has limited resources at its disposal to achieve its assigned mission, it must allocate its resources wisely. CNRC's most important resource is its recruiters. Accordingly, the locations to which they are assigned must be carefully chosen; they should be assigned to areas where the active-duty enlistment[4] supply is the greatest.

[1] U.S. Department of the Navy, *Human Capital Strategy 2007: Building and Managing the Total Naval Force*, Office of the Secretary of the Navy, 5.

[2] Navy Recruiting Command Public Affairs Office, "2008 Facts and Stats," Navy Recruiting Command, http://www.cnrc.navy.mil/PAO/facts_stats.htm (accessed May 9, 2008).

[3] Navy Recruiting Command Public Affairs Office, "2008 Facts and Stats," Navy Recruiting Command, http://www.cnrc.navy.mil/PAO/facts_stats.htm (accessed May 9, 2008).

[4] Active duty enlistment supply is specifically referred to here because the FY08 demand is 39,000, comprising 75% of total FY08 demand. Traditionally, recruiters and NRSs have been allocated based on the active duty enlisted market.

As many econometric studies have shown, many factors affect the supply of people willing to enlist into the United States Navy. Some relevant factors are:

- Number of recruiters;
- Advertising;
- Unemployment rate;
- Per capita income;
- Military pay; and
- Population.

Econometric theory postulates that localities should produce varying levels of enlistment supply in accordance with their values of the factors listed above. Accordingly, many models have been produced and tested with actual data. The results have generally supported the theory, but many variables have appeared to be extraneous.

One purpose of this study is to explore how proximity to a military installation, a factor that has not been addressed in previous studies, affects local enlistment supply. People who reside near a military facility likely view the military differently than those who do not. This proximity is likely to affect the amount of information available about military service, the perceived risks associated with military service, and the perceived rewards afforded to those in military service.

A second purpose of this study is to explore the how various measures of high school quality affect local enlistment supply. Since high schools provide the largest single source for navy enlisted applicants, the quality of the high schools in an NRS's territory should affect the quantity of high-quality applicants that join the Navy.

Additionally, this study seeks to explore the variables used in previous econometric studies as well as proximity to a military installation to produce a model to predict enlistment supply at the NRS level. This study seeks to develop the model with the highest predictive power. The results of this study should directly benefit CNRC in allocation of recruiting resources and in generating realistic expectations for NRS production.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

A. ENLISTMENT SUPPLY AT THE LOCAL MARKET LEVEL

Hogan et al. developed regression models to analyze enlistment supply at the zip code level. The authors used data from Army and Navy databases from FY 1994 to FY 1997 and data from the 1990 Census to estimate the parameters. For the recruiting station level model, the study used a log-log regression model and found that increasing the number of recruiters in a station was associated with an increase in the number of high-quality enlistments. Additionally, the authors posed several areas for future research. Specifically, they asked, "Does proximity to a military installation affect recruiting? If so, does it matter which service is located at the installation?"[5]

B. ANALYZING THE ASSIGNMENT OF ENLISTED RECRUITING GOAL SHARES VIA THE NAVY'S ENLISTED GOALING AND FORECASTING MODEL

In his thesis, Hojnowski provided an in-depth explanation of CNRC's active-duty enlisted goaling model, discussed the goaling model's performance versus actual production, and proposed adjustments to the model that may improve the accuracy of its predictions. The author explained that CNRC's goaling model is an econometric supply model that uses a fixed-effect, autoregressive estimator to predict high-quality male navy enlistments at the NRD level. Specific data sources used in estimating CNRC's model are not discussed. According to the author, some of the most important factors, based only on coefficients, are the number of recruiters, high quality-male population, low-quality male population, unemployment rate, and relative earnings.[6]

---

[5] Paul F. Hogan et al., "Enlistment Supply at the Local Market Level," (Technical Report NPS-SM-00-004, Naval Postgraduate School), 9-33.

[6] Ronald A. Hojnowski, "Analyzing the Assignment of Enlisted Recruiting Goal Shares Via the Navy's Enlisted Goaling and Forecasting Model," (Master's Thesis, Naval Postgraduate School, 2005), 35-40.

C. ALLOCATION OF RECRUITING RESOURCES ACROSS NAVY RECRUITING STATIONS AND METROPOLITAN AREAS

In their thesis, Jarosz and Stephens developed regression models for contract production at both recruiting station and metropolitan levels to assist CNRC in allocating recruiter resources. The parameters of the model were estimated using FY 1995 through FY 1997 data from U.S. Army Recruiting Command, CNRC, the Bureau of Labor Statistics, and the Census Bureau. The authors estimated both linear and log-log models at the NRS level using regression. The study explored how many different variables affected high-quality Navy enlistments. Of particular note, it showed that increasing the number of recruiters in a NRS generally led to an increase in high-quality enlistments.[7]

D. A STATISTICAL ESTIMATION OF NAVY ENLISTMENT SUPPLY MODELS USING ZIP CODE LEVEL DATA

Hostetler's thesis used Census Bureau data as well as FY 1996 zip code level data supplied by CNRC from its Standardized Territory Evaluation and Analysis for Management (STEAM) database to predict new contract production. The author developed a linear model and used the data to estimate the coefficients. This study also explored the collinearity among the independent variables, since many of the population demographics proved to be highly collinear. The author concluded that recruiter presence, a factor derived from number of recruiters in a station and the station's associated population, was the most important factor in predicting new contracts.[8]

[7] Suzanne K. Jarosz and Elisabeth S. Stephens, "Allocation of Recruiting Resources Across Navy Recruiting Stations and Metropolitan Areas," (Master's Thesis, Naval Postgraduate School, 1999), 2-54.

[8] David L. Hostetler, "A Statistical Estimation of Navy Enlistment Supply Models Using ZIP Code Level Data," (Master's Thesis, Naval Postgraduate School, 1998), 13-33.

E. PREDICTING THE NUMBER OF POTENTIAL MILITARY RECRUITS OVER THE NEXT TEN YEARS WITH APPLICATION TO RECRUITER PLACEMENT

Britton's thesis used zip code level data supplied by CNRC from July 2001 to June 2007 and DMDC data from FY 1998 to FY 2006 to evaluate CNRC's recruiter placement. This study assigned Navy applicants to categories based on demographics. The study then determined the ratio of applicants to general population for each demographic category. These ratios were then applied to each zip code to predict how many applicants it should have produced. By comparing the predicted value to the actual value, the study was able to estimate the propensity of a given zip code's population to enlist into the Navy. Through the same techniques, the study was able to provide propensities for various aggregates, e.g., NRS, NRD, Regional, and National.[9]

[9] Donald L. Britton, "Predicting the Number of Potential Military Recruits Over the Next Ten years with Application to Recruiter Placement," (Master's Thesis, Naval Postgraduate School, September 2007), xv-15.

THIS PAGE INTENTIONALLY LEFT BLANK

III. MODELS

Three general model types were chosen to predict high-quality male Navy enlistments: ordinary least squares multiple linear regression models, regression trees, and neural networks. These models were chosen because they allow for numerical response variables and because they are available in many software packages. All models used in this study were built using the data-mining software package SPSS Clementine 11.1. The names of the models below were based on the model selected in Clementine along with the particular settings chosen. For all models in this study, the response variable was the number of males with Armed Forces Qualification Test (AFQT) scores fifty or higher who entered the Navy's DEP. The full set of predictor variables was provided to each modeling tool as input. The algorithms for each model chose the variables to retain and their relative importance. Table 3 in Appendix A lists all variables derived as described in Appendix B. All variables in Appendix A, Table 3, except for the number of high-quality males (MU), recruiting station identification number (RSID), and the year, were used as predictor variables. The data from years 2002-2005 were used for training and the data from 2006 were used for testing. This training set and test set were chosen to provide a prediction environment similar to one that CNRC would experience in predicting the following fiscal year's enlistment supply.

A. REGRESSION[10]

Regression models comprised four of the five models in this study's literature review and are often used to gain insight into relationships between response variables and predictor variables. All of the regression models that

[10] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining, *Introduction to Linear Regression Analysis* (New York: John Wiley and Sons, Inc., 2004), 6.

were used in this study were ordinary least squares multiple linear regression models. The four models differ by the variable selection process, as described below.

1. Enter[11]

Here, “enter” refers to the variable selection option in the regression model in Clementine and designates that all variables were used—this is the full model.

2. Forwards[12]

The forward selection model begins with the simplest model—no predictor variables. Predictor variables are then added to the model if they improve the model. The predictor that improves the model the best is added in each step. The minimum requirement for variable entry was that the p-value associated with the F-statistic must be greater than 0.05.

3. Backwards[13]

Backwards elimination begins with the full model and then selects variables to remove at each step by removing the variable with the least statistical significance. This continues until all the variables that remain are statistically significant. Variable selection was complete when no variables remaining in the model had associated p-values greater than 0.1.

4. Stepwise[14]

Stepwise selection is the same as the forward selection model except that in each step, after a variable is added, the model is reevaluated to see if any variable currently in the model has become statistically insignificant. If so, one of

[11] Montgomery et al, 302.

[12] Montgomery et al, 310.

[13] Montgomery et al, 312.

[14] Montgomery et al, 314.

them is removed. Variables became candidates for removal when their associated p-values became greater than 0.1. The minimum requirement for variable entry was that the associated p-value must be greater than 0.05.

B. TREE[15]

Trees are produced by dividing data into sets that are more similar than they were before being divided. The splits that produced the highest degree of similarity are chosen. This process continues on each set that is produced until some stopping criteria are met. The three models below are differentiated by the number of splits allowed at each node and the method for finding optimal splits.

1. Tree: C&RT[16]

C&RT stands for Classification and Regression Tree. Since the response variable used in this study is continuous, this model specifically used the regression tree component of the algorithm. C&RT allows only binary splits at each node. For this model, default settings were used. Specifically, the Gini impurity method was used to measure similarity, the minimum change in impurity allowed was 0.0001, only five levels below the root were allowed, and the pruning option was selected.

2. Tree: CHAID[17]

CHAID stands for Chi-Squared Automatic Interaction Detector. CHAID is similar to C&RT, but it allows more than one split at each node (i.e., the tree is not required to be binary). Clementine 11.1 default settings were used.

[15] Montgomery et al, 516.

[16] Montgomery et al, 516.

[17] *Clementine 11.1 Algorithms Guide* (United States of America: Integral Solutions Limited, 2007), 44.

3. Tree: Exhaustive CHAID[18]

Exhaustive CHAID is a modification to the CHAID algorithm that overcomes CHAID's occasional inability to find the optimal split[19]. This results in longer computation times. Clementine 11.1 default settings were used.

C. NEURAL NETWORK[20]

A neural network is a statistical model that employs a network of interconnected weighting factors to convert input values into output values. The model uses the various predictor values and their associated response values to adjust the weights until the predicted response values are similar to the actual response values. Neural networks can provide good predictions, but do not normally provide insight into relationships between predictor variables and response variables. The six neural network models used in this study were the basic algorithms selectable in Clementine: Quick, Dynamic, Prune, Multiple, RBFN, and Exhaustive Prune. The quick method creates a network structure based on rules of thumb and data characteristics. The dynamic method creates a network structure similar to the quick method, but it allows the structure to be modified during training. The prune method begins with a large network and removes weak connections during training. The multiple method creates multiple networks with different structures and trains each of them. The model with the lowest error is selected. RBFN stands for Radial Basis Function Network and this method uses a clustering algorithm to aid in developing the network and to determine weighting factors. The exhaustive prune method is similar to the prune method but uses more thorough search techniques to find the weakest connections. For each neural network model, Clementine 11.1 default settings were used.[21]

---

[18] *Clementine 11.1 Algorithms Guide*, 44.

[19] Details on the weaknesses and how they are overcome can be found in *Clementine 11.1 Algorithms Guide* on pages 44-52.

[20] Montgomery et al, 518.

[21] *Clementine 11.1 Algorithms Guide*, 1-13.

IV. RESULTS AND CONCLUSIONS

A. RESULTS

1. Prediction Models

After the models were built using the 2002-2005 data, they were then used to predict the number of high-quality male enlistments for 2006. These predicted values were then compared to the actual values for 2006. The mean absolute errors were then calculated for each model. Three of the regression variable selection algorithms, forward selection, backwards elimination, and stepwise regression, produced the same mean absolute errors. This was due to the fact that in this study each method of variable selection technique ultimately resulted in the same model.[22] Surprisingly, the neural network models consistently outperformed both the regression and tree models. This was not expected at the outset of this study, as regression models have traditionally been used to predict enlistment supply. The regression models performed almost as well as the neural network models and much better than the tree models. Table 1 contains the mean absolute errors for each model. More detailed results are contained in Appendix C.

[22] These variable selection techniques, in general, may or may not lead to different models.

General Model Type	Specific Model Type	Mean Absolute Error (number of high-quality males joining the Navy's DEP per NRS per year)
Neural Network	Quick	6.194
	Dynamic	6.126
	Prune	5.845
	Multiple	6.010
	RBFN	7.141
	Exhaustive Prune	5.944
Tree	C & RT	6.765
	CHAID	6.901
	CHAID Exhaustive	6.734
Regression	Enter	6.141
	Forwards	6.142
	Backwards	6.142
	Stepwise	6.142

Table 1. Mean Absolute Error Table

2. Variables

a. Importance

In order to determine which factors were the most important, each variable was ranked, if possible, as to the order of importance in each model. For the regression models determined by the forward selection technique and the stepwise techniques, the rank was determined by the entering order. The results of the full regression model and regression model determined by the backwards elimination technique were not used in ranking the variables.[23] For the trees, the rank assigned was according to the level for which the variable was used as a split. For neural network models, the relative importance level as determined by Clementine's sensitivity analysis served as the rank.[24] Except for the full regression model and the regression model determined by the backwards elimination technique, the average rank was computed across all models and used for comparison.

By this metric, the number of recruiters per station, the 17-19 year old male population, the number of houses, and the veteran percentage proved to be the most important variables. The results of this analysis are listed in Table 2.

Table 2 also shows which variables were not included in some of the models. There were only eight variables that were retained by all of the models in this study. Those were the four listed above along with the percentage of students receiving subsidized lunches, the land area, the proximity to Navy installations factor, and per capita income.

Of the four most important variables, only the number of housing units (House) was surprising. The number of recruiters per station, the number of 17- to 19-year-old males, and the veteran population percentage were all used in the in various models covered in the literature review. Initially, the number of housing units may not appear to be a logical predictor of high-quality Navy enlistments. However, the number of housing units may serve as an interaction term between population and income level. This may be a worthwhile area for future research.

Each factor, student-to-teacher ratio, subsidized lunches, Promoting Power, and proximity to Navy installations, was identified by at least one measure to be important in predicting high-quality male Navy enlistment supply. Subsidized lunches and Navy installation proximity proved significant by

---

[23] The full model and the backwards elimination model do not provide any real insight into variable importance beyond whether their inclusion is statistically significant. Further, the backwards model contains the same variables as the forwards and stepwise models. Therefore, it is not necessary for the inclusion/ exclusion analysis either.

[24] *Clementine 11.1 Algorithms Guide*, 11.

being ranked sixth and eighth in importance, as calculated by the average rank metric, and by being included in every model. The variable selection process for regression chose Promoting Power scores and student-to-teacher ratios as important in predicting high-quality Navy enlistment supply. Demonstrating the importance of these variables allowed for meaningful exploration of their relationships in the next section.

Variables\Model Type	Reg:F	Reg:S	CRT	CHAID	E CHAID	NN:Q	NN:D	NN:P	NN:M	NN:R	NN:EP	**Avg Rank**
RPS	1	1	1	1	1	1	2	2	2	1	3	**1.45**
M17	2	2	2	2	2	2	1	1	1	3	1	**1.73**
House	9	9	5	2	2	6	4	10	3	7	2	**5.36**
Vet Per	3	3	4	3	3	7	11	9	8	9	4	**5.82**
M20 *	6	6	6	6	6	12	7	4	5	4	5	**6.09**
SubLunch	4	4	3	2	2	5	10	11	9	13	10	**6.64**
LArea	10	10	4	3	3	4	9	3	7	12	8	**6.64**
Navy_P_D	5	5	6	4	4	17	3	5	14	14	15	**8.36**
M17_25 **	18	18	5	2	2	3	6	12	6	15	6	**8.45**
PerCapB	8	8	5	3	6	8	8	14	10	18	9	**8.82**
M20_25 *	11	11	6	6	6	11	12	6	12	10	16	**9.73**
Warea *	12	12	6	4	4	9	17	17	4	11	12	**9.82**
UmempB *	13	13	6	6	6	10	15	15	13	8	11	**10.55**
Avg Dis ***	18	18	6	3	3	18	5	8	18	6	18	**11.00**
HS ***	18	18	6	7	6	14	16	16	11	2	7	**11.00**
Non_Navy_P_D ***	18	18	6	4	6	15	14	7	17	5	14	**11.27**
STRatio *	7	7	6	6	6	13	18	18	16	16	17	**11.82**
Score *	14	14	6	6	6	16	13	13	15	17	13	**12.09**

* Not included in at least one tree model

** Not included in at least one regression model

*** Not included in at least one tree and one regression model

Reg:F: Regression model determined by the forward selection technique.

Reg:S: Regression model determined by the forward selection technique.

CRT: Tree model determined by the C & RT algorithm.

CHAID: Tree model determined by the CHAID algorithm.

E CHAID: Treem model determined by the Exhaustive CHAID algorithm.

NN:Q: Neural Network model determined by the Quick algorithm.

NN:D: Neural Network model determined by the Dynamic algorithm.

NN:P: Neural Network model determined by the Prune algorithm.

NN:M: Neural Network model determined by the Multiple algorithm.

NN:R: Neural Network model determined by the RBFN algorithm.

NN:EP: Neural Network model determined by the Exhaustive Prune algorithm.

Table 2. Variables Ranked by Importance

b. Relationships

This study's regression model, as determined by the stepwise selection technique, was used to evaluate relationships between predictor variables and the response variables. In general, the relationships established in this study between predictor variables and the response variable appeared logical and were in agreement with those in the literature review. The relationships of the military proximity variable and school quality variables with high-quality male Navy enlistments are detailed below.

(1) Student-to-teacher Ratio. The student-to-teacher ratio (STRatio) was statistically significant in this model and had a positive regression coefficient. Thus, as the student-to-teacher ratio increases, the predicted number of high-quality male Navy enlistments tends to also increase for an NRS. At first, this result seemed rather counter-intuitive because high student-to-teacher ratios are often associated with lower-quality schools. However, a strong relationship may exist between very small class sizes and very high college enrollment rates. Assuming this increase in college enrollment results in fewer enlistments into the Navy, an increased student-to-teacher ratio would then be serving as a proxy for reduced college enrollment rates. Further examination as to why an increase in student-to-teacher ratios results in an increase in high-quality male Navy enlistments is an area for future research.

(2) Promoting Power Score. The Promoting Power scores (Score), indicators of high school graduation rates, were statistically significant in this model and had a negative regression coefficient. Thus, as the Promoting Power score increases, the predicted number of high-quality male Navy enlistments tends to decrease for an NRS. As with the result for student-to-teacher ratios, this result initially appears to be unexpected. Higher graduation rates have generally been associated with higher school quality and, therefore, should yield more high-quality enlistments. Again, very high graduation rates could be indicative of very high college enrollment rates causing a decrease in the number of Navy enlistments.

(3) Subsidized Lunches. The percentage of students receiving subsidized lunches (SubLunch) was statistically significant in this model and had a negative regression coefficient. Thus, as the percentage of students receiving subsidized lunches increases, the predicted number of high-quality male Navy enlistments tends to decrease for an NRS. Here, if subsidized lunches are a true indicator of the quality of education, then this result seems reasonable. Since subsidized lunches are directly based off of income level, one might also expect that increasing the percentage of students receiving subsidized lunches and the associated decrease in the local civilian pay to military pay ratio might increase the number of high-quality enlistments. Of the metrics related to high school, the percentage of students receiving subsidized lunches ranked as most important and provided the expected relationship between high school quality and the number of high-quality male Navy enlistments.

(4) Proximity to Navy Installations. As expected, the ratio of Navy installation personnel to the distance between the Navy installation and the NRS (Navy_P_D_largest) was statistically significant in this model and had a positive regression coefficient. Thus, as the personnel to proximity ratio increases, the predicted number of high-quality male Navy enlistments tends to also increase for an NRS.

B. CONCLUSIONS

The purpose of this study was to build predictive models, to explore the relationship between military installation proximity and high-quality male Navy enlistments, and to explore the relationships between various high school quality factors and high-quality male Navy enlistments. Through comparing the mean absolute errors between predicted and actual results, the neural network models outperformed both regression and tree models. The study also showed that Navy installation proximity and various measures of high school quality are significant in predicting the number of high-quality male Navy enlistments. Furthermore, the study verified that the number of high-quality male Navy enlistments was larger for an NRS when the distance between an NRS and a Navy installation was small and when the population of a nearby Navy installation was large. The number of high-quality male Navy enlistments was higher for an NRS when the NRS's territory contained public high schools with higher student-to-teacher ratios, lower graduation rates (as demonstrated by Promoting Power scores), and fewer students receiving subsidized lunches.

This study indicated that CNRC may be able to develop better enlistment production forecasts and associated recruiter assignment models by using neural network models to supplement their regression based models. Also, the accuracy of their models may be improved by incorporating proximities to military installations as well as measures of high-school quality.

Future studies may increase the fidelity of the predictions as well as the relationships between the predictor variables and response variable by improving on the data set used in this study. Specifically, zip code level data with annual measurements should be used for all records and fields. Additionally, such factors as distances between NRSs and Military Entrance Processing Stations (MEPSs), distances between NRSs and the NRD headquarters, types and numbers of colleges and universities, and the number of Junior Reserve Officers Training Corps (JROTC) units should be explored in enlistment supply models in future studies. Finally, further research should be conducted in order to validate the relationships between the number of housing units and the number of high-quality male Navy enlistments and between high school quality and the number of high-quality male Navy enlistments and to further elucidate the underlying reasons for those relationships.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. VARIABLE DESCRIPTIONS

Variable Name	Variable Description	Data Source
AvgDis	Average distance from centroid of an NRS's zip code to the centroid of each zip code that that NRS's area of responsibility.	Calculation
HS	Number of high schools in an NRS's area of responsibility.	Census 2000
House	Number of housing units in an NRS's area of responsibility.	Census 2000
LArea	Land area in square miles in an NRS's area of responsibility.	Census 2000
M17	Number of males age 17-19 in an NRS's area of responsibility.	Woods and Poole
M17_25	Number of males age 17-19 within zip codes whose centroid is within 25 miles of an NRS's zip code's centroid.	Woods and Poole
M20	Number of males age 20-24 in an NRS's area of responsibility.	Woods and Poole
M20_25	Number of males age 20-24 within zip codes whose centroid is within 25 miles of an NRS's zip code's centroid.	Woods and Poole
MU	Number of males with an AFQT score 50 or higher who joined the Navy's DEP.	CNRC
Navy_P_D_largest	The largest value of (number of people)/(distance + 1) representing an NRS's proximity to a Navy installation and the distance from that installation based on population categories.	Base Status Report
Non_Navy_P_D_largest	The largest value of (number of people)/(distance + 1) representing an NRS's proximity to a Non-Navy installation and the distance from that installation based on population categories.	Base Status Report
PerCapB	Per capita income within an NRS's area of responsibility.	Department of Labor
RPS	Average number of recruiters assigned to an NRS.	CNRC
RSID	Recruiting station identification number assigned to an NRS.	CNRC
Score	"Promoting Power" score representing the public high school graduation rate in an NRS's area of responsibility.	Alliance for Excellent Education
STRatio	Student to teacher ratio for public high schools in an NRS's area of responsibility.	Department of Education
SubLunch	Percentage of public high school students receiving reduced or free lunches in an NRS's area of responsibility.	Department of Education
UnempB	The unemployment rate in an NRS's area of responsibility.	Department of Labor
VetPer	Percentage of the population in an NRS's area of responsibility that are military veterans.	Census 2000
WArea	Water area in square miles in an NRS's area of responsibility.	Census 2000
Year	Fiscal Year from which data was produced.	Census 2000

Table 3. Description of Variables

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. DATA

A. DATA SOURCES

1. CNRC

CNRC provided several sources of data to Britton for use in his thesis[25]. This data, with considerable amounts of pre-processing performed on it, was made available for follow-on theses.

a. Woods and Poole Economics, Inc.

CNRC provided Britton population data from Woods and Poole Economics, Inc., "an independent firm that specializes in long-term county economic and demographic projections."[26] This data contained population counts categorized by age, gender, race, and education level for each county and zip code in the United States. There were three datasets provided: documented residence status, undocumented residence status, and total population. Each dataset contained 29,583,180 records with 29 fields.

b. Zip Code and FIPS Mapping to Recruiting Stations

CNRC provided Britton a file containing 41,400 zip codes mapped to their associated Federal Information Processing Standards (FIPS) code and local NRS. Each NRS is identified by a unique recruiting station identification number (RSID). This file was important, as zip codes, FIPS codes, and RSIDs were used as keys to merge files and aggregate data.

---

[25] Britton.

[26] Woods & Poole Economics, "Woods & Poole Economics, Washington, D.C.: County Forecasts to 2030," Woods & Poole Economics, http://www.woodsandpoole.com/ (accessed May 25, 2008).

c. Latitude and Longitude for Each Zip Code

CNRC provided Britton a flat file containing the latitude and longitude of the centroid for 41,520 zip codes. This file was used for calculating distances between zip codes.

d. Navy Recruiting Station Manning Levels

CNRC provided Britton a file containing specific recruiter information such as report date, transfer date, and the recruiting station assignment. Through processing by Britton, this file provided average annual recruiter manning levels for 1051 NRSs identified by RSID for 2001 through 2007.

e. Census Data

CNRC provided Britton a file containing data from the 2000 Census. The data provided included land area in square miles, water area in square miles, the number of public high schools, and the number of houses for each zip code.

2. Defense Manpower Data Center (DMDC)

DMDC, the Department of Defense's source for human resource information, provided to Britton a data set consisting of military service applicants from FY 1998 through FY 2006. This file contained applicant information such as Armed Forces Qualification Test (AFQT) scores, gender, age, race, home of record zip code, Delayed Entry Program (DEP) entry date, and DEP service for every component (active, reserve, and guard) of each service (Air Force, Army, Coast Guard, Marine Corps, and Navy). This data set contained 18 fields and 4,296,409 records.

3. U.S. Census Bureau

Three data files containing unemployment information, veteran populations, and per capita income were downloaded from the U.S. Census Bureau's Download Center.[27] The data came from Summary File 3 of the 2000 U.S. Census. Each data file contained nearly 32,000 zip code tabulation areas (ZCTA) which approximate the geographic delivery areas for U.S. Postal Service zip codes. The number of ZCTAs available in each Census file is about 10,000 fewer than the number of zip codes provided in CNRCs zip code file. This is due to the fact that the Census Bureau assigns three-digit ZCTAs to large contiguous areas for which it does not have five-digit zip code information available. The per capita income comprised five fields: zip codes, per capita incomes, and three geographic identifiers. The veteran population file consisted of 27 fields broken down by sex and age. The file containing unemployment information was arranged in 19 fields and consisted of population counts and the number of unemployed persons for various demographic segments.

4. U.S. Department of Commerce

Annual county-level per capita income and population files were provided via download from the U.S. Department of Commerce's Bureau of Economic Analysis Website.[28] The data consisted of per capita incomes and populations for 3133 counties for each year from 2002 through 2006. Counties were identified via FIPS codes.

[27] U.S. Census Bureau, "U.S. Census Bureau: American Fact Finder," U.S. Census Bureau, http://factfinder.census.gov/servlet/DCGeoSelectServlet?ds_name=DEC_2000_SF3_U (accessed May 19, 2008).

[28] U.S. Department of Commerce Bureau of Economic Analysis, "Bureau of Economic Analysis: Regional Economic Accounts," U.S. Department of Commerce, http://www.bea.gov/regional/reis/ (accessed May 19, 3008).

5. U.S. Department of Labor

Annual county-level unemployment files were provided via download from the U.S. Department of Labor's Bureau of Labor Statistics Website.[29] Five files were downloaded, one for each year from 2002-2006. Each file contained the number of people in the labor force, number of people employed, and number of people unemployed for 3224 counties. Each of the files used identical formats.

6. U.S. Department of Defense

An Excel file containing military installation data was extracted from an Adobe Portable Document Format (pdf) copy of the Department of Defense's Base Structure Report (BSR): Fiscal Year 2003 Baseline.[30] The data consisted of Total Replacement Value (PRV), total number of personnel authorized for the site, primary component owner of the site, and site zip code for 1,132 military sites.

7. U.S. Department of Education

Files containing information about U.S. public high schools were downloaded from the U.S. Department of Education's National Center for Education Statistics Website.[31] Fifty-one data files (one for each state plus Washington, D.C.) were downloaded; each contained 37 fields covering 18,180 high schools. The data was gathered from the 2005-2006 school year. Among

[29] U.S. Department of Labor Bureau of Labor Statistics, "U.S. Department of Labor Bureau of Labor Statistics: Local Area Unemployment Statistics," U.S. Department of Labor, http://www.bls.gov/lau/ (accessed May 7, 2008).

[30] U.S. Department of Defense, *Department of Defense, Base Structure Report (A Summary of DoD's Real Property Inventory): Fiscal Year 2003 Baseline,* Office of the Deputy Under Secretary of Defense, Installations and Environment.

[31] U.S. Department of Education Institute of Education Sciences National Center for Education Statistics, "IES National Center for Education Statistics: Search for Public Schools," U.S. Department of Education, http://nces.ed.gov/ccd/schoolsearch/ (accessed on May 19, 2008).

the fields were the number of students, the number of teachers, the number of students receiving free lunches, and the number of students receiving reduced lunches.

8. Alliance for Excellent Education

Due to the multiple ways that public high school graduation rates are calculated, a consistent indicator of graduation rates is necessary. Researchers at Johns Hopkins University have created an indicator for high school graduation rates called "Promoting Power." This statistic compares the number of seniors in a high school to the number of ninth-graders enrolled in the high school three years earlier. Fifty-one files, one for each state and Washington, D.C., were downloaded from Alliance for Excellent Education's Website.[32] Each file contained "Promoting Power" scores for public high schools for 2004, 2005, and 2006. A total of 15,208 records were contained in the downloaded files.

B. DATA PREPARATION

1. Individual Data File Preparation

The data files were modified to produce fields (columns) for each desired variable and to produce records (rows) for each NRS and year combination. Most files required only minor modification, mapping zip codes to NRSs and then summing up the fields for each aggregated NRS and zip code combination. The county level data required FIPS codes as keys to be mapped to NRSs. Some counties, however, were mapped to multiple NRSs potentially introducing error into the data. Additionally, some data sources did not contain data for each year covered in this study, so imputation was necessary. The data that did not

[32] Alliance for Excellent Education, "High Schools in the United States: How Does Your Local High School Measure Up?" Alliance for Excellent Education, http://www.all4ed.org/about_the_crisis/schools/state_and_local_info/promotingpower (accessed on May 13, 2008).

conform to time periods or the geographic boundaries of this study, but were used due to availability, and the manipulations performed on them are listed below.

a. U.S. Census Bureau

U.S. Census Bureau data was provided for each zip code, but only for a single year. Per capita income, veteran population, and unemployment data were from the 2000 Census was used as a constant value for 2002, 2003, etc. An average weighted by population was used to aggregate per capita income from a zip code level to an NRS level.

b. U.S. Department of Commerce and U.S. Department of Labor

The per capita income data provided by the U.S. Department of Commerce and the unemployment data provided by the U.S. Department of Labor were provided for each year, but they were provided only at the county level. During aggregation, the unemployment data were summed, and a weighted average was taken of per capita income. However, since county boundaries and NRS boundaries do not always coincide, it was not possible to equitably divide and weight the data during aggregation to the NRS level.[33] This causes some NRSs to potentially have overly inflated or deflated per capita incomes and unemployment rates.

c. U.S. Department of Defense

Records from the BSR data file with empty zip code or total personnel fields were removed from the file. Latitude and longitude fields were then merged with the military installation file with zip codes used as the merge key. The distance between each NRS and each military installation was then calculated using latitudes and longitudes of the associated zip codes. The

[33] Approximately 26% of the counties mapped to multiple NRSs.

closest Navy installation and non-Navy installation were identified for five different installation sizes based on number of authorized personnel: greater than 100, 500, 1000, 2500, or 5000.

In order to evaluate the factors of both proximity and number of personnel at the same time without masking effects from larger installations that might only be more distant by a few miles, a new factor was created. The following calculation was performed for each combination of installation type (Navy and non-Navy) and installation size for each NRS:

$$\frac{\text{number of authorized personnel}}{(\text{distance in miles from NRS + 1 mile})}.$$[34]

The largest value for a Navy installation and for a non-Navy installation were retained with the NRS and denoted as Navy_P_D_largest and Non_Navy_P_D_largest respectively.

d. Alliance for Excellent Education

The Alliance for Excellent Education provided “Promoting Power” scores and zip codes for each high school, but it did not provide associated high school populations. Since there were no unique identifiers to pair up the 15,208 public high schools with their populations, an unweighted average of the “Promoting Power” scores was calculated during aggregation to NRS levels. Additionally, only scores for 2004, 2005, and 2006 were provided and some of those scores were missing. Any missing values of the 2004-2006 scores were filled with the average value of the provided scores for that high school. The 2002 and 2003 scores also had to be imputed. The scores for 2002 and 2003 were filled with the 2004 score.

[34] One mile was added to each distance in order to prevent division by zero for those NRSs that were located in the same ZIP code as the military installation.

2. Data File Merging

In order to efficiently import data into data analysis software, a single file was created containing all pertinent fields from the individual files and records for each NRS and year. NRS RSIDs and years were used as the merge keys. Only records with keys common to all data sets were used for estimating parameters in this study.

C. DATA AUDITING

After the data files were merged into a single file, an audit of the data was performed. The audit showed that the fields containing number of migrant students and the percentage of students receiving subsidized lunches contained several missing values. The migrant student field contained a significant number of missing values and was removed from the data file, but all records were retained. Of the 4,848 records, 224 contained missing values for percentage of students receiving subsidized lunches. Analyzing the distribution of missing values for subsidized lunches indicated that they were not randomly distributed. Most of the missing values were in records form NRSs in Arizona, Nevada, Texas, Tennessee, and Wisconsin. The concentration of missing values in specific geographic locations did cause some concern. Since the percentage of students receiving subsidized lunches was an important variable to be studied and over 95% of the records contained valid values, this field was retained. The 224 records containing missing values were removed from the data set.

APPENDIX C. MODELING RESULTS

A. REGRESSION: ENTER

Variable Name	Beta Coefficients	P-value
(constant)	0.48800	0.848
M17	0.00740	0.000
M20	-0.00613	0.005
M17_25	-0.00159	0.326
M20_25	0.00343	0.139
PerCapB	-0.00013	0.000
UnempB	0.33300	0.005
VetPer	0.46700	0.000
Score	-0.03200	0.052
STRatio	0.22000	0.000
SubLunch	-0.13000	0.000
RPS	3.63100	0.000
HS	-0.00348	0.817
AvgDis	-0.00405	0.639
House	0.00002	0.000
LArea	0.00018	0.000
WArea	-0.00446	0.086
Navy_P_D_largest	0.00100	0.000
Non_Navy_P_D_largest	-0.00001	0.960

Table 4. Regression: Enter Model Results

B. REGRESSION: FORWARDS, BACKWARDS, AND STEPWISE

Variable Name	Beta Coefficients	P-value
(constant)	0.55900	0.825
RPS	3.62800	0.000
M17	0.00595	0.000
VetPer	0.46500	0.000
SubLunch	-0.13000	0.000
Navy_P_D_largest	0.00101	0.000
M20	-0.00416	0.000
STRatio	0.22200	0.000
PerCapB	-0.00013	0.000
House	0.00002	0.000
LArea	0.00050	0.000
M20_25	0.00127	0.021
WArea	-0.00559	0.001
UnempB	0.32600	0.006
Score	-0.03270	0.043

Table 5. Regression: Variable Selection Models Results

C. TREE: C&RT

- RPS <= 3.540 [Ave: 21.515, Effect: -5.221] (2,114)
  - M17 <= 3,089 [Ave: 19.421, Effect: -2.094] (1,754)
    - RPS <= 2.375 [Ave: 16.867, Effect: -2.554] (927)
      - M17 <= 1758.500 [Ave: 14.98, Effect: -1.888] (537)
        - House <= 38,335 [Ave: 11.093, Effect: -3.886] ⇨ 11.093 (107)
        - House > 38,335 [Ave: 15.947, Effect: 0.967] ⇨ 15.947 (430)
      - M17 > 1758.500 [Ave: 19.467, Effect: 2.599] ⇨ 19.467 (390)
    - RPS > 2.375 [Ave: 22.284, Effect: 2.863] (827)
      - SubLunch <= 55.588 [Ave: 22.992, Effect: 0.708] (744)
        - M17 <= 1,672 [Ave: 20.733, Effect: -2.259] ⇨ 20.733 (221)
        - M17 > 1,672 [Ave: 23.946, Effect: 0.955] ⇨ 23.946 (523)
      - SubLunch > 55.588 [Ave: 15.94, Effect: -6.344] ⇨ 15.94 (83)
  - M17 > 3,089 [Ave: 31.717, Effect: 10.202] (360)
    - M17 <= 4,196 [Ave: 28.621, Effect: -3.096] (253)
      - VetPer <= 13.713 [Ave: 26.724, Effect: -1.897] ⇨ 26.724 (170)
      - VetPer > 13.713 [Ave: 32.506, Effect: 3.885] ⇨ 32.506 (83)
    - M17 > 4,196 [Ave: 39.037, Effect: 7.321] ⇨ 39.037 (107)
- RPS > 3.540 [Ave: 33.605, Effect: 6.869] (1,607)
  - M17 <= 3,034 [Ave: 30.337, Effect: -3.269] (924)
    - SubLunch <= 40.016 [Ave: 32.74, Effect: 2.404] (562)
      - RPS <= 5.040 [Ave: 30.796, Effect: -1.944] (432)
        - M17_25 <= 1,706 [Ave: 27.497, Effect: -3.3] ⇨ 27.497 (149)
        - M17_25 > 1,706 [Ave: 32.534, Effect: 1.737] ⇨ 32.534 (283)
      - RPS > 5.040 [Ave: 39.2, Effect: 6.46] (130)
        - PerCapB <= 31378.083 [Ave: 34.889, Effect: -4.311] ⇨ 34.889 (72)
        - PerCapB > 31378.083 [Ave: 44.552, Effect: 5.352] ⇨ 44.552 (58)
    - SubLunch > 40.016 [Ave: 26.605, Effect: -3.732] ⇨ 26.605 (362)
  - M17 > 3,034 [Ave: 38.028, Effect: 4.422] (683)
    - RPS <= 4.460 [Ave: 34.644, Effect: -3.384] (323)
      - M17 <= 4735.500 [Ave: 33.08, Effect: -1.564] (274)
        - LArea <= 452.348 [Ave: 29.32, Effect: -3.761] ⇨ 29.32 (97)
        - LArea > 452.348 [Ave: 35.141, Effect: 2.061] ⇨ 35.141 (177)
      - M17 > 4735.500 [Ave: 43.388, Effect: 8.744] ⇨ 43.388 (49)
    - RPS > 4.460 [Ave: 41.064, Effect: 3.036] (360)
      - LArea <= 735.998 [Ave: 37.562, Effect: -3.502] (162)
        - House <= 214381.500 [Ave: 35.61, Effect: -1.952] ⇨ 35.61 (118)
        - House > 214381.500 [Ave: 42.795, Effect: 5.234] ⇨ 42.795 (44)
      - LArea > 735.998 [Ave: 43.929, Effect: 2.865] ⇨ 43.929 (198)

Table 6. Tree: C&RT Model Results

D. TREE: CHAID

- RPS <= 1.670 [Ave: 15.573, Effect: -11.163] (361)
  - M17_25 <= 875 [Ave: 11.802, Effect: -3.771] ⇨ 11.802 (101)
  - M17_25 > 875 and M17_25 <= 2,269 [Ave: 15.286, Effect: -0.287] ⇨ 15.286 (220)
  - M17_25 > 2,269 [Ave: 26.675, Effect: 11.102] ⇨ 26.675 (40)
- RPS > 1.670 and RPS <= 2 [Ave: 18.926, Effect: -7.811] (376)
  - House <= 50,768 [Ave: 14.657, Effect: -4.268] ⇨ 14.657 (70)
  - House > 50,768 and House <= 84,811 [Ave: 17.338, Effect: -1.588] (157)
    - PerCapB <= 31503.864 [Ave: 16.43, Effect: -0.908] ⇨ 16.43 (114)
    - PerCapB > 31503.864 [Ave: 19.744, Effect: 2.407] ⇨ 19.744 (43)
  - House > 84,811 and House <= 127,036 [Ave: 19.652, Effect: 0.727] ⇨ 19.652 (92)
  - House > 127,036 [Ave: 27.368, Effect: 8.443] ⇨ 27.368 (57)
- RPS > 2 and RPS <= 2.420 [Ave: 20.877, Effect: -5.86] (358)
  - House <= 50,768 [Ave: 16.5, Effect: -4.377] ⇨ 16.5 (54)
  - House > 50,768 and House <= 127,036 [Ave: 19.746, Effect: -1.131] (244)
    - SubLunch <= 28.319 [Ave: 20.33, Effect: 0.584] (109)
      - Non_Navy_P_D_largest <= 113.409 [Ave: 18.444, Effect: -1.886] ⇨ 18.444 (63)
      - Non_Navy_P_D_largest > 113.409 [Ave: 22.913, Effect: 2.583] ⇨ 22.913 (46)
    - SubLunch > 28.319 and SubLunch <= 35.349 [Ave: 22.123, Effect: 2.377] ⇨ 22.123 (57)
    - SubLunch > 35.349 [Ave: 17.192, Effect: -2.554] ⇨ 17.192 (78)
  - House > 127,036 [Ave: 29.417, Effect: 8.54] ⇨ 29.417 (60)
- RPS > 2.420 and RPS <= 2.920 [Ave: 23.22, Effect: -3.517] (410)
  - M17 <= 1,809 [Ave: 19.255, Effect: -3.965] (157)
    - SubLunch <= 45.265 [Ave: 21.114, Effect: 1.859] ⇨ 21.114 (114)
    - SubLunch > 45.265 [Ave: 14.326, Effect: -4.929] ⇨ 14.326 (43)
  - M17 > 1,809 and M17 <= 3,429 [Ave: 23.18, Effect: -0.039] ⇨ 23.18 (205)
  - M17 > 3,429 [Ave: 36.354, Effect: 13.135] ⇨ 36.354 (48)
- RPS > 2.920 and RPS <= 3.170 [Ave: 24.929, Effect: -1.808] (324)
  - M17 <= 2,073 [Ave: 21.773, Effect: -3.156] (132)
    - AvgDis <= 12.624 [Ave: 24.073, Effect: 2.3] ⇨ 24.073 (55)
    - AvgDis > 12.624 [Ave: 20.13, Effect: -1.643] ⇨ 20.13 (77)
  - M17 > 2,073 and M17 <= 2,977 [Ave: 24.705, Effect: -0.224] ⇨ 24.705 (112)
  - M17 > 2,977 [Ave: 30.45, Effect: 5.521] ⇨ 30.45 (80)
- RPS > 3.170 and RPS <= 3.670 [Ave: 27.612, Effect: 0.876] (436)
  - M17 <= 2,073 [Ave: 23.207, Effect: -4.405] (135)
    - SubLunch <= 45.265 [Ave: 25.656, Effect: 2.449] ⇨ 25.656 (96)
    - SubLunch > 45.265 [Ave: 17.179, Effect: -6.028] ⇨ 17.179 (39)
  - M17 > 2,073 and M17 <= 3,429 [Ave: 26.886, Effect: -0.727] (201)
    - LArea <= 1196.854 [Ave: 25.462, Effect: -1.424] ⇨ 25.462 (104)
    - LArea > 1196.854 and LArea <= 2292.707 [Ave: 32.366, Effect: 5.48] ⇨ 32.366 (41)
    - LArea > 2292.707 [Ave: 25.518, Effect: -1.368] ⇨ 25.518 (56)
  - M17 > 3,429 and M17 <= 4,121 [Ave: 32.774, Effect: 5.161] ⇨ 32.774 (53)
  - M17 > 4,121 [Ave: 37.553, Effect: 9.941] ⇨ 37.553 (47)

- RPS > 3.670 and RPS <= 4 [Ave: 29.898, Effect: 3.161] (353)
  - M17 <= 3,429 [Ave: 27.977, Effect: -1.921] (266)
    - SubLunch <= 39.491 [Ave: 29.903, Effect: 1.926] (165)
      - Navy_P_D_largest <= 37.396 [Ave: 32.548, Effect: 2.645] ⇨ 32.548 (62)
      - Navy_P_D_largest > 37.396 and Navy_P_D_largest <= 156.955 [Ave: 26.475, Effect: -3.428] ⇨ 26.475 (61)
      - Navy_P_D_largest > 156.955 [Ave: 30.976, Effect: 1.073] ⇨ 30.976 (42)
    - SubLunch > 39.491 [Ave: 24.832, Effect: -3.146] ⇨ 24.832 (101)
  - M17 > 3,429 and M17 <= 4,121 [Ave: 31.864, Effect: 1.966] ⇨ 31.864 (44)
  - M17 > 4,121 [Ave: 39.767, Effect: 9.869] ⇨ 39.767 (43)
- RPS > 4 and RPS <= 4.420 [Ave: 32.814, Effect: 6.077] (333)
  - M17 <= 2,641 [Ave: 29.429, Effect: -3.385] (154)
    - SubLunch <= 39.491 [Ave: 31.235, Effect: 1.806] (98)
      - M17_25 <= 1,699 [Ave: 27.304, Effect: -3.93] ⇨ 27.304 (46)
      - M17_25 > 1,699 [Ave: 34.712, Effect: 3.477] ⇨ 34.712 (52)
    - SubLunch > 39.491 [Ave: 26.268, Effect: -3.161] ⇨ 26.268 (56)
  - M17 > 2,641 and M17 <= 4,121 [Ave: 34.04, Effect: 1.227] (124)
    - VetPer <= 14.020 [Ave: 31.446, Effect: -2.595] ⇨ 31.446 (83)
    - VetPer > 14.020 [Ave: 39.293, Effect: 5.252] ⇨ 39.293 (41)
  - M17 > 4,121 [Ave: 39.527, Effect: 6.713] ⇨ 39.527 (55)
- RPS > 4.420 and RPS <= 5.170 [Ave: 34.709, Effect: 7.973] (406)
  - M17 <= 2,977 [Ave: 30.042, Effect: -4.668] (216)
    - SubLunch <= 45.265 [Ave: 31.756, Effect: 1.715] (160)
      - WArea <= 0.759 [Ave: 30.439, Effect: -1.317] ⇨ 30.439 (41)
      - WArea > 0.759 and WArea <= 4.808 [Ave: 36.732, Effect: 4.976] ⇨ 36.732 (56)
      - WArea > 4.808 [Ave: 28.19, Effect: -3.566] ⇨ 28.19 (63)
    - SubLunch > 45.265 [Ave: 25.143, Effect: -4.899] ⇨ 25.143 (56)
  - M17 > 2,977 and M17 <= 4,121 [Ave: 37.439, Effect: 2.729] (114)
    - VetPer <= 12.757 [Ave: 32.135, Effect: -5.304] ⇨ 32.135 (52)
    - VetPer > 12.757 [Ave: 41.887, Effect: 4.449] ⇨ 41.887 (62)
  - M17 > 4,121 [Ave: 43.882, Effect: 9.172] ⇨ 43.882 (76)
- RPS > 5.170 [Ave: 38.643, Effect: 11.906] (364)
  - SubLunch <= 39.491 [Ave: 41.774, Effect: 3.131] (221)
    - M17 <= 2,641 [Ave: 37.884, Effect: -3.89] ⇨ 37.884 (69)
    - M17 > 2,641 [Ave: 43.539, Effect: 1.766] ⇨ 43.539 (152)
  - SubLunch > 39.491 and SubLunch <= 53.741 [Ave: 36.267, Effect: -2.376] (90)
    - M17 <= 2,977 [Ave: 32.091, Effect: -4.176] ⇨ 32.091 (44)
    - M17 > 2,977 [Ave: 40.261, Effect: 3.994] ⇨ 40.261 (46)
  - SubLunch > 53.741 [Ave: 29.623, Effect: -9.02] ⇨ 29.623 (53)

Table 7. Tree: CHAID Model Results

E. TREE: EXHAUSTIVE CHAID

- RPS <= 1.670 [Ave: 15.573, Effect: -11.163] (361)
  - M17_25 <= 875 [Ave: 11.802, Effect: -3.771] ⇨ 11.802 (101)
  - M17_25 > 875 and M17_25 <= 2,269 [Ave: 15.286, Effect: -0.287] ⇨ 15.286 (220)
  - M17_25 > 2,269 [Ave: 26.675, Effect: 11.102] ⇨ 26.675 (40)
- RPS > 1.670 and RPS <= 2 [Ave: 18.926, Effect: -7.811] (376)
  - House <= 50,768 [Ave: 14.657, Effect: -4.268] ⇨ 14.657 (70)
  - House > 50,768 and House <= 84,811 [Ave: 17.338, Effect: -1.588] ⇨ 17.338 (157)
  - House > 84,811 and House <= 127,036 [Ave: 19.652, Effect: 0.727] ⇨ 19.652 (92)
  - House > 127,036 [Ave: 27.368, Effect: 8.443] ⇨ 27.368 (57)
- RPS > 2 and RPS <= 2.420 [Ave: 20.877, Effect: -5.86] (358)
  - M17 <= 1,260 [Ave: 15.778, Effect: -5.099] ⇨ 15.778 (63)
  - M17 > 1,260 and M17 <= 2,977 [Ave: 19.906, Effect: -0.971] (235)
    - SubLunch <= 28.319 [Ave: 20.385, Effect: 0.479] ⇨ 20.385 (109)
    - SubLunch > 28.319 and SubLunch <= 35.349 [Ave: 22.38, Effect: 2.474] ⇨ 22.38 (50)
    - SubLunch > 35.349 [Ave: 17.592, Effect: -2.314] ⇨ 17.592 (76)
  - M17 > 2,977 [Ave: 30.033, Effect: 9.156] ⇨ 30.033 (60)
- RPS > 2.420 and RPS <= 2.920 [Ave: 23.22, Effect: -3.517] (410)
  - M17 <= 1,809 [Ave: 19.255, Effect: -3.965] (157)
    - SubLunch <= 45.265 [Ave: 21.114, Effect: 1.859] ⇨ 21.114 (114)
    - SubLunch > 45.265 [Ave: 14.326, Effect: -4.929] ⇨ 14.326 (43)
  - M17 > 1,809 and M17 <= 3,429 [Ave: 23.18, Effect: -0.039] ⇨ 23.18 (205)
  - M17 > 3,429 [Ave: 36.354, Effect: 13.135] ⇨ 36.354 (48)
- RPS > 2.920 and RPS <= 3.170 [Ave: 24.929, Effect: -1.808] (324)
  - M17 <= 2,073 [Ave: 21.773, Effect: -3.156] (132)
    - AvgDis <= 12.624 [Ave: 24.073, Effect: 2.3] ⇨ 24.073 (55)
    - AvgDis > 12.624 [Ave: 20.13, Effect: -1.643] ⇨ 20.13 (77)
  - M17 > 2,073 and M17 <= 2,977 [Ave: 24.705, Effect: -0.224] ⇨ 24.705 (112)
  - M17 > 2,977 [Ave: 30.45, Effect: 5.521] ⇨ 30.45 (80)
- RPS > 3.170 and RPS <= 3.670 [Ave: 27.612, Effect: 0.876] (436)
  - M17 <= 2,073 [Ave: 23.207, Effect: -4.405] (135)
    - SubLunch <= 45.265 [Ave: 25.656, Effect: 2.449] ⇨ 25.656 (96)
    - SubLunch > 45.265 [Ave: 17.179, Effect: -6.028] ⇨ 17.179 (39)
  - M17 > 2,073 and M17 <= 3,429 [Ave: 26.886, Effect: -0.727] (201)
    - LArea <= 1196.854 [Ave: 25.462, Effect: -1.424] ⇨ 25.462 (104)
    - LArea > 1196.854 and LArea <= 2292.707 [Ave: 32.366, Effect: 5.48] ⇨ 32.366 (41)
    - LArea > 2292.707 [Ave: 25.518, Effect: -1.368] ⇨ 25.518 (56)
  - M17 > 3,429 and M17 <= 4,121 [Ave: 32.774, Effect: 5.161] ⇨ 32.774 (53)
  - M17 > 4,121 [Ave: 37.553, Effect: 9.941] ⇨ 37.553 (47)

- RPS > 3.670 and RPS <= 4 [Ave: 29.898, Effect: 3.161] (353)
  - M17 <= 3,429 [Ave: 27.977, Effect: -1.921] (266)
    - SubLunch <= 39.491 [Ave: 29.903, Effect: 1.926] (165)
      - Navy_P_D_largest <= 37.396 [Ave: 32.548, Effect: 2.645] ⇨ 32.548 (62)
      - Navy_P_D_largest > 37.396 and Navy_P_D_largest <= 156.955 [Ave: 26.475, Effect: -3.428] ⇨ 26.475 (61)
      - Navy_P_D_largest > 156.955 [Ave: 30.976, Effect: 1.073] ⇨ 30.976 (42)
    - SubLunch > 39.491 [Ave: 24.832, Effect: -3.146] ⇨ 24.832 (101)
  - M17 > 3,429 and M17 <= 4,121 [Ave: 31.864, Effect: 1.966] ⇨ 31.864 (44)
  - M17 > 4,121 [Ave: 39.767, Effect: 9.869] ⇨ 39.767 (43)
- RPS > 4 and RPS <= 4.420 [Ave: 32.814, Effect: 6.077] (333)
  - M17 <= 2,641 [Ave: 29.429, Effect: -3.385] ⇨ 29.429 (154)
  - M17 > 2,641 and M17 <= 4,121 [Ave: 34.04, Effect: 1.227] (124)
    - VetPer <= 14.020 [Ave: 31.446, Effect: -2.595] ⇨ 31.446 (83)
    - VetPer > 14.020 [Ave: 39.293, Effect: 5.252] ⇨ 39.293 (41)
  - M17 > 4,121 [Ave: 39.527, Effect: 6.713] ⇨ 39.527 (55)
- RPS > 4.420 and RPS <= 5.170 [Ave: 34.709, Effect: 7.973] (406)
  - M17 <= 2,977 [Ave: 30.042, Effect: -4.668] (216)
    - SubLunch <= 45.265 [Ave: 31.756, Effect: 1.715] (160)
      - WArea <= 0.759 [Ave: 30.439, Effect: -1.317] ⇨ 30.439 (41)
      - WArea > 0.759 and WArea <= 4.808 [Ave: 36.732, Effect: 4.976] ⇨ 36.732 (56)
      - WArea > 4.808 [Ave: 28.19, Effect: -3.566] ⇨ 28.19 (63)
    - SubLunch > 45.265 [Ave: 25.143, Effect: -4.899] ⇨ 25.143 (56)
  - M17 > 2,977 and M17 <= 4,121 [Ave: 37.439, Effect: 2.729] (114)
    - VetPer <= 12.757 [Ave: 32.135, Effect: -5.304] ⇨ 32.135 (52)
    - VetPer > 12.757 [Ave: 41.887, Effect: 4.449] ⇨ 41.887 (62)
  - M17 > 4,121 [Ave: 43.882, Effect: 9.172] ⇨ 43.882 (76)
- RPS > 5.170 [Ave: 38.643, Effect: 11.906] (364)
  - SubLunch <= 39.491 [Ave: 41.774, Effect: 3.131] (221)
    - M17 <= 2,641 [Ave: 37.884, Effect: -3.89] ⇨ 37.884 (69)
    - M17 > 2,641 [Ave: 43.539, Effect: 1.766] ⇨ 43.539 (152)
  - SubLunch > 39.491 and SubLunch <= 53.741 [Ave: 36.267, Effect: -2.376] ⇨ 36.267 (90)
  - SubLunch > 53.741 [Ave: 29.623, Effect: -9.02] ⇨ 29.623 (53)

Table 8. Tree: Exhaustive CHAID Model Results

F. NEURAL NETWORK: QUICK

Relative Importance of Inputs

RPS	0.359158
M17	0.24773
M17_25	0.153058
LArea	0.133628
SubLunch	0.130363
House	0.126843
VetPer	0.105586
PerCapB	0.0898765
WArea	0.0635858
UnempB	0.057381
M20_25	0.0522404
M20	0.0516538
STRatio	0.0499411
HS	0.0155088
Non_Navy_P_D_largest	0.0135273
Score	0.0112047
Navy_P_D_largest	0.0101779
AvgDis	0.00315254

Table 9. Neural Network: Quick Model Results

G. NEURAL NETWORK: DYNAMIC

Relative Importance of Inputs

M17	0.398039
RPS	0.39398
Navy_P_D_largest	0.189243
House	0.188676
AvgDis	0.160411
M17_25	0.151287
M20	0.144424
PerCapB	0.141423
LArea	0.141266
SubLunch	0.135766
VetPer	0.131647
M20_25	0.118415
Score	0.107053
Non_Navy_P_D_largest	0.0801814
UnempB	0.068962
HS	0.0686954
WArea	0.0662357
STRatio	0.0648299

Table 10. Neural Network: Dynamic Model Results

H. NEURAL NETWORK: PRUNE

Relative Importance of Inputs

M17	0.477286
RPS	0.325413
LArea	0.267921
M20	0.264005
Navy_P_D_largest	0.255575
M20_25	0.173654
Non_Navy_P_D_largest	0.167141
AvgDis	0.157786
VetPer	0.148353
House	0.143121
SubLunch	0.142351
M17_25	0.131819
Score	0.115589
PerCapB	0.112932
UnempB	0.108297
HS	0.0989722
WArea	0.0760732
STRatio	0.070351

Table 11. Neural Network: Prune Model Results

I. NEURAL NETWORK: MULTIPLE

Relative Importance of Inputs

M17	0.420731
RPS	0.380898
House	0.185358
WArea	0.173822
M20	0.148958
M17_25	0.145339
LArea	0.127175
VetPer	0.109843
SubLunch	0.109161
PerCapB	0.0985052
HS	0.086079
M20_25	0.0803487
UnempB	0.0683105
Navy_P_D_largest	0.0589769
Score	0.0517954
STRatio	0.0463956
Non_Navy_P_D_largest	0.0385041
AvgDis	0.0337244

Table 12. Neural Network: Multiple Model Results

J. NEURAL NETWORK: RPFN

Relative Importance of Inputs

RPS	0.150526
HS	0.142804
M17	0.118415
M20	0.118106
Non_Navy_P_D_largest	0.0968185
AvgDis	0.0965332
House	0.0793094
UnempB	0.0748115
VetPer	0.0741514
M20_25	0.072758
WArea	0.0684081
LArea	0.0675429
SubLunch	0.0636026
Navy_P_D_largest	0.0625797
M17_25	0.0605929
STRatio	0.0597612
Score	0.058998
PerCapB	0.0581132

Table 13. Neural Network: RPFN Model Results

K. NEURAL NETWORK: EXHAUSTIVE PRUNE

Relative Importance of Inputs

M17	0.51627
House	0.45997
RPS	0.31179
VetPer	0.249134
M20	0.234695
M17_25	0.185663
HS	0.145329
LArea	0.13953
PerCapB	0.132086
SubLunch	0.108877
UnempB	0.100217
WArea	0.0961913
Score	0.0854292
Non_Navy_P_D_largest	0.0842722
Navy_P_D_largest	0.0756569
M20_25	0.0705913
STRatio	0.0579261
AvgDis	0.0424783

Table 14. Neural Network: Exhaustive Prune Model Results

LIST OF REFERENCES

Alliance for Excellent Education. "High Schools in the United States: How You're your Local High School Measure Up?" Alliance for Excellent Education. http://www.all4ed.org/about_the_crisis/schools/state_and_local_info/promotingpower (accessed on May 13, 2008).

Britton, Donald L. "Predicting the Number of Potential Military Recruits over the Next Ten Years with Application to Recruiter Placement." Master's Thesis, Naval Postgraduate School, 2007.

Clementine 11.1 Algorithms Guide. United States of America: Integral Solutions Limited, 2007.

Hogan, Paul F., Jared Hughes, Stephen Mehay, and Michael D. Cook. "Enlistment Supply at the Local Market Level." Technical Report NPS-SM-00-0004, Naval Postgraduate School, 2000.

Hojnowski, Ronald A. "Analyzing the Assignment of Enlisted Recruiting Goal Shares Via the Navy's Enlisted Goaling and Forecasting Model." Master's Thesis, Naval Postgraduate School, 2005.

Hostetler, David L. "A Statistical Estimation of Navy Enlistment Supply Models Using Zip Code Level Data." Master's Thesis, Naval Postgraduate School, 1998.

Jarosz, Suzanne K., and Elisabeth S. Stephens. "Allocation of Recruiting Resources Across Navy Recruiting Stations and Metropolitan Areas." Master's Thesis, Naval Postgraduate School 1999.

Montgomery, Douglas C, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis.* New York: John Wiley and Sons, Inc., 2004.

Navy Recruiting Command Public Affairs Office. "2008 Facts and Stats." Navy Recruiting Command. http://www.cnrc.navy.mil/PAO/facts_stats.htm (accessed on May 9, 2008).

U.S. Census Bureau. "U.S. Census Bureau: American Fact Finder." U.S. Census Bureau. http://factfinder.census.gov/servlet/DCGeoSelectServlet?ds_name=DEC_2000_SF3_U (accessed May 19, 2008).

U.S. Department of Commerce Bureau of Economic Analysis. "Bureau of Economic Analysis: Regional Economic Accounts." U.S. Department of Commerce. http://www.bea.gov/regional/reis/ (accessed May 19, 3008).

U.S. Department of Defense. *Department of Defense, Base Structure Report (A Summary of DoD's Real Property Inventory): Fiscal Year 2003 Baseline.* Prepared by the Office of the Deputy Under Secretary of Defense, Installations and Environment.

U.S. Department of Education Institute of Education Sciences National Center for Education Statistics. "IES National Center for Education Statistics: Search for Public Schools." U.S. Department of Education. http://nces.ed.gov/ccd/schoolsearch/ (accessed on May 19, 2008).

U.S. Department of Labor Bureau of Labor Statistics. "U.S. Department of Labor Bureau of Labor Statistics: Local Area Unemployment Statistics." U.S. Department of Labor. http://www.bls.gov/lau/ (accessed May 7, 2008).

U.S. Department of the Navy. *Human Capital Strategy 2007: Building and Managing the Total Naval Force.* Prepared by the Office of the Secretary of the Navy, Manpower and Reserve Affairs.

Woods & Poole Economics. "Woods & Poole Economics, Washington, D.C.: County Forecasts to 2030." Woods & Poole Economics. http://www.woodsandpoole.com/ (accessed May 25, 2008).

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Professor Samuel E. Buttrey
   Naval Postgraduate School
   Monterey, California

4. CDR David L. Schiffman
   Naval Postgraduate School
   Monterey, California

5. Mr. Wayne Wagner
   Arlington Navy Annex Department of Navy
   Washington, District of Columbia

6. Mr. John Noble
   Navy Recruiting Command
   Millington, Tennessee

7. CAPT Maureen Cahill
   Naval Postgraduate School
   Monterey, California

8. Director, Temasek Defence Systems Institute
   National University of Singapore
   Singapore

9. LCDR Claude M. McRoberts
   Arlington Navy Annex Department of Navy
   Washington, District of Columbia